

Subscribe (Full Service) Register (Limited Service, Free) Login

Search: The ACM Digital Library The Guide

(cluster < near/4> concepts) frequency occurrences



# THE ACH DIGITAL LIBRAR

Feedback Report a problem Satisfaction survey

Terms used <u>cluster near/4 concepts frequency occurrences</u>

Found 26,328 of 154,226

Sort results relevance bν

Save results to a Binder Search Tips

Try an Advanced Search Try this search in The ACM Guide

Display results

expanded form

Open results in a new window

Result page: 1 2 3 4 5 6 7 8 9 10

Best 200 shown

Results 1 - 20 of 200

Relevance scale 🖵 📟

1 Concepts and effectiveness of the cover-coefficient-based clustering methodology for text databases

Fazli Can, Esen A. Ozkarahan

December 1990 ACM Transactions on Database Systems (TODS), Volume 15 Issue 4

Full text available: R pdf(2.74 MB)

Additional Information: full citation, abstract, references, citings, index terms, review

A new algorithm for document clustering is introduced. The base concept of the algorithm, the cover coefficient (CC) concept, provides a means of estimating the number of clusters within a document database and related indexing and clustering analytically. The CC concept is used also to identify the cluster seeds and to form clusters with these seeds. It is shown that the complexity of the clustering process is very low. The retrieval experiments show that the information-retrieval effectiv ...

Keywords: cluster validity, clustering-indexing relationships, cover coefficient, decoupling coefficient, document retrieval, retrieval effectiveness

A framework for effective retrieval

C. T. Yu, W. Meng, S. Park

June 1989 ACM Transactions on Database Systems (TODS), Volume 14 Issue 2

Full text available: R pdf(1,56 MB)

Additional Information: full citation, abstract, references, citings, index terms, review

The aim of an effective retrieval system is to yield high recall and precision (retrieval effectiveness). The nonbinary independence model, which takes into consideration the number of occurrences of terms in documents, is introduced. It is shown to be optimal under the assumption that terms are independent. It is verified by experiments to yield significant improvement over the binary independence model. The nonbinary model is extended to normalized vectors and is applicable to more genera ...

Concepts of the cover coefficient-based clustering methodology

Fazli Can, Esen A. Ozkarahan

June 1985 Proceedings of the 8th annual international ACM SIGIR conference on Research and development in information retrieval

Full text available: ndf(745.86 KB) Additional Information: full citation, abstract, references, citings

Document clustering has several unresolved problems. Among them are high time and



space complexity, difficulty of determining similarity thresholds, order dependence, nonuniform document distribution in clusters, and arbitrariness in determination of various cluster intiators. To overcome these problems to some degree, the cover coefficient based clustering methodology has been introduced. The concepts used in this methodology have created certain new concepts, relationships, and me ...

4 Special issue on word sense disambiguation: Introduction to the special issue on word sense disambiguation: the state of the art



Nancy Ide, Jean Véronis

March 1998 Computational Linguistics, Volume 24 Issue 1



Additional Information: full citation, references, citings

Deriving concept hierarchies from text

Mark Sanderson, Bruce Croft

August 1999 Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval

Full text available: pdf(100.05 KB) Additional Information: full citation, references, citings, index terms

Keywords: concept hierarchy, multi-document summary, subsumption, term co-occurence

6 Poster papers: CVS: a Correlation-Verification based Smoothing technique on information retrieval and term clustering



Christina Yip Chung, Bin Chen

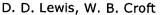
July 2002 Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining

Full text available: pdf(634.38 KB) Additional Information: full citation, abstract, references, index terms

As information volume in enterprise systems and in the Web grows rapidly, how to accurately retrieve information is an important research area. Several corpus based smoothing techniques have been proposed to address the data sparsity and synonym problems faced by information retrieval systems. Such smoothing techniques are often unable to discover and utilize the correlations among terms. We propose CVS, a Correlation-Verification based Smoothing method, that considers co-occurrence information i ...

Keywords: information retrieval, query expansion, smoothing, term clustering, text mining

Term clustering of syntactic phrases





Full text available: pdf(1.62 MB)

Additional Information: full citation, abstract, references, citings, index

Term clustering and syntactic phrase formation are methods for transforming natural language text. Both have had only mixed success as strategies for improving the quality of text representations for document retrieval. Since the strengths of these methods are complementary, we have explored combining them to produce superior representations. In this paper we discuss our implementation of a syntactic phrase generator, as well as our preliminary experiments with producing phrase clusters. Th ...

8 Information access and retrieval: Multiple related document summary and navigation using concept hierarchies for mobile clients



D. L. Chan, R. W. P. Luk, W. K. Mak, H. V. Leong, E. K. S. Ho, Q. Lu March 2002 Proceedings of the 2002 ACM symposium on Applied computing

Full text available: ndf(660.35 KB) Additional Information: full citation, abstract, references, index terms

Mobile clients have limited display and navigation capabilities. To browse a set of documents, an intuitive method is to navigate through concept hierarchies. To reduce semantic loading for each term that represents the concepts and the cognitive loading of users due to the limited display, similar documents are grouped together before concept hierarchies are constructed for each document group. Since the concept hierarchies only represent the salient concepts in the documents, term extraction i ...

Keywords: browsing, concept hierarchy, information access, mobile agent, mobile computing, navigation, summarization

Papers: Concept clustering and knowledge integration from a children's dictionary Caroline Barrière, Fred Popowich



August 1996 Proceedings of the 16th conference on Computational linquistics - Volume

Full text available: pdf(568.40 KB) Additional Information: full citation, abstract, references

Knowledge structure called Concept Clustering Knowledge Graphs (CCKGs) are introduced along with a process for their construction from a machine readable dictionary. CCKGs contain multiple concepts interrelated through multiple semantic relations together forming a semantic cluster represented by a conceptual graph. The knowledge acquisition is performed on a children's first dictionary. The concepts involved are general and typical of a daily life conversation. A collection of conceptual cluste ...

10 The concept of dynamic analysis

Thoms Bell

October 1999 ACM SIGSOFT Software Engineering Notes, Proceedings of the 7th European software engineering conference held jointly with the 7th ACM SIGSOFT international symposium on Foundations of software engineering, Volume 24 Issue 6

Full text available: R pdf(1.37 MB)

Additional Information: full citation, abstract, references, citings, index terms

Dynamic analysis is the analysis of the properties of a running program. In this paper, we explore two new dynamic analyses based on program profiling: Frequency Spectrum Analysis. We show how analyzing the frequencies of program entities in a single execution can help programmers to decompose a program, identify related computations, and find computations related to specific input and output characteristics of a program. Cover ...

11 Fast detection of communication patterns in distributed executions



Thomas Kunz, Michiel F. H. Seuren

November 1997 Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research

Full text available: pdf(4.21 MB)

Additional Information: full citation, abstract, references, index terms

Understanding distributed applications is a tedious and difficult task. Visualizations based on process-time diagrams are often used to obtain a better understanding of the execution of the application. The visualization tool we use is Poet, an event tracer developed at the

University of Waterloo. However, these diagrams are often very complex and do not provide the user with the desired overview of the application. In our experience, such tools display repeated occurrences of non-trivial commun ...

12 Semantic indexing for a complete subject discipline

Yi-Ming Chung, Qin He, Kevin Powell, Bruce Schatz

August 1999 Proceedings of the fourth ACM conference on Digital libraries

Full text available: pdf(256.74 KB) Additional Information: full citation, references, index terms

Keywords: MEDLINE, MEDSPACE, concept space, interspace, medical informatics, scalable semantics, semantic indexing, semantic retrieval

13 Web clustering: Inferring hierarchical descriptions

Eric Glover, David M. Pennock, Steve Lawrence, Robert Krovetz

November 2002 Proceedings of the eleventh international conference on Information and knowledge management

Full text available: pdf(239.32 KB)

Additional Information: full citation, abstract, references, citings, index

We create a statistical model for inferring hierarchical term relationships about a topic, given only a small set of example web pages on the topic, without prior knowledge of any hierarchical information. The model can utilize either the full text of the pages in the cluster or the context of links to the pages. To support the model, we use "ground truth" data taken from the category labels in the Open Directory. We show that the model accurately separates terms in the following classes: sel ...

Keywords: cluster naming, feature selection, hierarchical relationships, statistical models, web analysis

14 Research track posters: Cluster-based concept invention for statistical relational learning

Alexandrin Popescul, Lyle H. Ungar

August 2004 Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining

Full text available: pdf(157.81 KB) Additional Information: full citation, abstract, references, index terms

We use clustering to derive new relations which augment database schema used in automatic generation of predictive features in statistical relational learning. Entities derived from clusters increase the expressivity of feature spaces by creating new first-class concepts which contribute to the creation of new features. For example, in CiteSeer, papers can be clustered based on words or citations giving "topics", and authors can be clustered based on documents they co-author giving "communities" ...

Keywords: clustering, feature generation, relational learning

15 Using multiple knowledge sources for word sense discrimination

Susan W. McRoy

March 1992 Computational Linguistics, Volume 18 Issue 1

Full text available: Additional Information: full citation, abstract, references, citings



This paper addresses the problem of how to identify the intended meaning of individual words in unrestricted texts, without necessarily having access to complete representations of sentences. To discriminate senses, an understander can consider a diversity of information, including syntactic tags, word frequencies, collocations, semantic context, rolerelated expectations, and syntactic restrictions. However, current approaches make use of only small subsets of this information. Here we will des ...

16 An error-based conceptual clustering method for providing approximate guery answers W. W. Chu, K. Chiang, C. Hsu, H. Yau

December 1996 Communications of the ACM

Full text available: pdf(351.78 KB) Additional Information: full citation, references, citings, index terms

17 On modeling of information retrieval concepts in vector spaces S. K.M. Wong, W. Ziarko, V. V. Raghavan, P. C.N. Wong June 1987 ACM Transactions on Database Systems (TODS), Volume 12 Issue 2

Full text available: mpdf(1.80 MB)

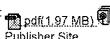
Additional Information: full citation, abstract, references, citings, index terms, review

The Vector Space Model (VSM) has been adopted in information retrieval as a means of coping with inexact representation of documents and queries, and the resulting difficulties in determining the relevance of a document relative to a given query. The major problem in employing this approach is that the explicit representation of term vectors is not known a priori. Consequently, earlier researchers made the assumption that the vectors corresponding to terms are pairwise orthogonal. Such an a ...

18 Special issue on word sense disambiguation: Automatic word sense discrimination Hinrich Schütze



March 1998 Computational Linguistics, Volume 24 Issue 1



Full text available: pdf(1.97 MB) Additional Information: full citation, abstract, references, citings

This paper presents context-group discrimination, a disambiguation algorithm based on clustering. Senses are interpreted as groups (or clusters) of similar contexts of the ambiguous word. Words, contexts, and senses are represented in Word Space, a highdimensional, real-valued space in which closeness corresponds to semantic similarity. Similarity in Word Space is based on second-order co-occurrence: two tokens (or contexts) of the ambiguous word are assigned to the same sense cluster if the wo ...

19 A survey of Web metrics

Devanshu Dhyani, Wee Keong Ng, Sourav S. Bhowmick December 2002 ACM Computing Surveys (CSUR), Volume 34 Issue 4

Full text available: pdf(289.28 KB)

Additional Information: full citation, abstract, references, citings, index terms

The unabated growth and increasing significance of the World Wide Web has resulted in a flurry of research activity to improve its capacity for serving information more effectively. But at the heart of these efforts lie implicit assumptions about "quality" and "usefulness" of Web resources and services. This observation points towards measurements and models that quantify various attributes of web sites. The science of measuring all aspects of information, especially its storage and retrieval or ...

Keywords: Information theoretic, PageRank, Web graph, Web metrics, Web page similarity, quality metrics

20 Web searching: WISE-cluster; clustering e-commerce search engines automatically November 2004 Proceedings of the 6th annual ACM international workshop on Web information and data management



Full text available: pdf(366.41 KB) Additional Information: full citation, abstract

Results 1 - 20 of 200

Result page: 1 2 3 4 5 6 7 8 9 10 next

The ACM Portal is published by the Association for Computing Machinery. Copyright @ 2005 ACM, Inc. Terms of Usage Privacy Policy Code of Ethics Contact Us

Useful downloads: Adobe Acrobat QuickTime Windows Media Player



Subscribe (Full Service) Register (Limited Service, Free) Login

Search: The ACM Digital Library C The Guide

(cluster < near/4> concepts) frequency occurrences

## THE ACM DIG TALLISKARY

Feedback Report a problem Satisfaction survey

Terms used <u>cluster near/4 concepts frequency occurrences</u>

Found 26,328 of 154,226

Sort results bγ

relevance

Save results to a Binder Search Tips

Try an Advanced Search Try this search in The ACM Guide

Display results

expanded form

Open results in a new window

Results 21 - 40 of 200

Result page: previous 1 2 3 4 5 6 7 8 9 10

Relevance scale 🔲 📟 📟

Best 200 shown

21 A parallel algorithm for record clustering Edward Omiecinski, Peter Scheuermann

December 1990 ACM Transactions on Database Systems (TODS), Volume 15 Issue 4

Full text available: ndf(1.82 MB)

Additional Information: full citation, abstract, references, citings, index terms, review

We present an efficient heuristic algorithm for record clustering that can run on a SIMD machine. We introduce the P-tree, and its associated numbering scheme, which in the split phase allows each processor independently to compute the unique cluster number of a record satisfying an arbitrary query. We show that by restricting ourselves in the merge phase to combining only sibling clusters, we obtain a parallel algorithm whose speedup ratio is optimal in the number of processors used. Final ...

22 Exploiting clustering and phrases for context-based information retrieval

Peter G. Anick, Shivakumar Vaithyanathan

July 1997 ACM SIGIR Forum, Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval.

Volume 31 Issue SI

Full text available: pdf(1.55 MB)

Additional Information: full citation, references, citings, index terms

23 Concept based query expansion

Yonggang Qiu, Hans-Peter Frei

July 1993 Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval

Full text available: pdf(1.05 MB)

Additional Information: full citation, abstract, references, citings, index terms

Query expansion methods have been studied for a long time - with debatable success in many instances. In this paper we present a probabilistic query expansion model based on a similarity thesaurus which was constructed automatically. A similarity thesaurus reflects domain knowledge about the particular collection from which it is constructed. We address the two important issues with query expansion: the selection and the weighting of additional search terms. In contrast to earlier methods, ...

24 Building efficient and effective metasearch engines Weiyi Meng, Clement Yu, King-Lup Liu

## March 2002 ACM Computing Surveys (CSUR), Volume 34 Issue 1

Full text available: pdf(416.07 KB)

Additional Information: full citation, abstract, references, citings, index

Frequently a user's information needs are stored in the databases of multiple search engines. It is inconvenient and inefficient for an ordinary user to invoke multiple search engines and identify useful documents from the returned results. To support unified access to multiple search engines, a metasearch engine can be constructed. When a metasearch engine receives a query from a user, it invokes the underlying search engines to retrieve useful information for the user. Metasearch engines have ...

**Keywords**: Collection fusion, distributed collection, distributed information retrieval, information resource discovery, metasearch

25 Information retrieval & extraction: N-gram cluster identification during empirical knowledge representation generation



Robin Collier

August 1994 Proceedings of the 15th conference on Computational linguistics - Volume

Full text available: pdf(467.30 KB) Additional Information: full citation, abstract, references

This paper presents an overview of current research concerning knowledge extraction from technical texts. In particular, the use of empirical techniques during the identification and generation of a semantic representation is considered. A key step is the discovery of useful n-grams and correlations between clusters of these n-grams.

Keywords: knowledge representation, language understanding, large text corpora

<sup>26</sup> Improving the effectiveness of information retrieval with local context analysis Jinxi Xu, W. Bruce Croft



January 2000 ACM Transactions on Information Systems (TOIS), Volume 18 Issue 1

Full text available: pdf(193.02 KB)

Additional Information: full citation, abstract, references, citings, index terms, review

Techniques for automatic query expansion have been extensively studied in information research as a means of addressing the word mismatch between queries and documents. These techniques can be categorized as either global or local. While global techniques rely on analysis of a whole collection to discover word relationships, local techniques emphasize analysis of the top-ranked documents retrieved for a query. While local techniques have shown to be more effective that global techniques in ...

Keywords: cooccurrence, document analysis, feedback, global techniques, information retrieval, local context analysis, local techniques

27 Modeling word occurrences for the compression of concordances





Full text available: pdf(630.99 KB)

Additional Information: full citation, abstract, references, index terms, review

An earlier paper developed a procedure for compressing concordances, assuming that all alements occurred independently. The models introduced in that paper are extended here to take the possiblity of clustering into account. The concordance is conceptualized as a set of bitmaps, in which the bit locations reporesent documents, and the one-bits represent the occurrence of given terms. Hidden Markov Models (HMM's) are used to describe the clustering of the one-bits. However, for computational ...

Keywords: classification of graph nodes, concordance organization, concordance storage, graph structure

28 Evaluation of an inference network-based retrieval model

Howard Turtle, W. Bruce Croft

July 1991 ACM Transactions on Information Systems (TOIS), Volume 9 Issue 3

Full text available: pdf(2.40 MB)

Additional Information: full citation, references, citings, index terms, review

Keywords: document retrieval, inference networks, network retrieval models

29 IR-KM-1 (information retrieval and knowledge management): text mining: Dynamic extraction topic descriptors and discriminators: towards automatic context-based topic



Ana Maguitman, David Leake, Thomas Reichherzer, Filippo Menczer

November 2004 Proceedings of the Thirteenth ACM conference on Information and knowledge management

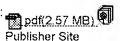
Full text available: pdf(253.70 KB) Additional Information: full citation, abstract, references, index terms

Effective knowledge management may require going beyond initial knowledge capture, to support decisions about how to extend previously-captured knowledge. Electronic <i>concept maps,</i> interlinked with other concept maps and multimedia resources, can provide rich <i>knowledge models</i> for human knowledge capture and sharing. This paper presents research on methods for supporting experts as they extend these knowledge models, by searching the Web for new context-relevant to ...

**Keywords:** acquisition tools, automatic topic search, concept mapping, context, information retrieval, knowledge, knowledge management

30 Word sense disambiguation using a second language monolingual corpus Ido Dagan, Alon Itai December 1994 Computational Linguistics, Volume 20 Issue 4





Full text available: pdf(2.57 MB) Additional Information: full citation, abstract, references, citings

This paper presents a new approach for resolving lexical ambiguities in one language using statistical data from a monolingual corpus of another language. This approach exploits the differences between mappings of words to senses in different languages. The paper concentrates on the problem of target word selection in machine translation, for which the approach is directly applicable. The presented algorithm identifies syntactic relations between words, using a source language parser, and maps t ...

31 Automatic abstracting and indexing—survey and recommendations H. P. Edmundson, R. E. Wyllys

May 1961 Communications of the ACM, Volume 4 Issue 5

Full text available: pdf(1.04 MB)

Additional Information: full citation, abstract, references, citings, index terms

In preparation for the widespread use of automatic scanners which will read documents and transmit their contents to other machines for analysis, this report presents a new concept in automatic analysis: the relative-frequency approach to measuring the significance of words, word groups, and sentences. The relative-frequency approach is discussed in detail, as is its application to problems of automatic indexing and automatic abstracting. Included in the report is a summary of automatic ana ...

32 Semantic Clustering of Index Terms



S. Kumar

October 1968 Journal of the ACM (JACM), Volume 15 Issue 4

Full text available: pdf(1.30 MB)

Additional Information: full citation, references, citings, index terms

33 HyPursuit: a hierarchical network search engine that exploits content-link hypertext clustering



Ron Weiss, Bienvenido Vélez, Mark A. Sheldon

March 1996 Proceedings of the the seventh ACM conference on Hypertext

Full text available: pdf(2.00 MB)

Additional Information: full citation, references, citings, index terms

34 Information Retrieval and Text Mining: A clustering algorithm for asymmetrically related data with applications to text mining



K. Krishna, Raghu Krishnapuram

October 2001 Proceedings of the tenth international conference on Information and knowledge management

Full text available: mpdf(593,39 KB)

Additional Information: full citation, abstract, references, citings, index terms

Clustering techniques find a collection of subsets of a data set such that the collection satisfies a criterion that is dependent on a relation defined on the data set. The underlying relation is traditionally assumed to be symmetric. However, there exist many practical scenarios where the underlying relation is asymmetric. One example of an asymmetric relation in text analysis is the inclusion relation, i.e., the inclusion of the meaning of a block of text in the meaning of another block. In th ...

35 An evaluation of phrasal and clustered representations on a text categorization task David D. Lewis



June 1992 Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval

Full text available: pdf(1.22 MB)

Additional Information: full citation, abstract, references, citings, index terms

Syntactic phrase indexing and term clustering have been widely explored as text representation techniques for text retrieval. In this paper we study the properties of phrasal and clustered indexing languages on a text categorization task, enabling us to study their properties in isolation from query interpretation issues. We show that optimal effectiveness occurs when using only a small proportion of the indexing terms available, and that effectiveness peaks at a higher feature set size and ...

36 The Logical Record Access Approach to Database Design

Toby J. Teorey, James P. Fry

June 1980 ACM Computing Surveys (CSUR), Volume 12 Issue 2

Full text available: pdf(2.81 MB) Additional Information: full citation, references, citings, index terms

37 Web mining, tools, and performance evaluation: Concept extraction and association from cancer literature



Yueyu Fu, Travis Bauer, Javed Mostafa, Mathew Palakal, Snehasis Mukhopadhyay November 2002 Proceedings of the 4th international workshop on Web information and data management

Full text available: pdf(272.50 KB)

Additional Information: full citation, abstract, references, citings, index terms

There is a large and growing body of web accessible biomedical literature. As this body of electronic literature grows, so does the possibility that document analysis techniques can be used to automatically extract useful biomedical information from them, particularly in the discovery of key concepts dealing with genes, proteins, drugs, and diseases and associations among these concepts. VCGS (Vocabulary Cluster Generating System) was designed to automatically extract and determine associations ...

Keywords: web data mining, web information extraction

38 The use of cluster hierarchies in hypertext information retrieval

D. B. Crouch, C. J. Crouch, G. Andreas

November 1989 Proceedings of the second annual ACM conference on Hypertext

Full text available: pdf(1.05 MB)

Additional Information: full citation, abstract, references, citings, index terms

The graph-traversal approach to hypertext information retrieval is a conceptualization of hypertext in which the structural aspects of the nodes are emphasized. A user navigates through such hypertext systems by evaluating the semantics associated with links between nodes as well as the information contained in nodes. [Fris88] In this paper we describe an hierarchical structure which effectively supports the graphical traversal of a document collection in a hypertext system ...

39 Information retrieval 1: Pruning long documents for distributed information retrieval Jie Lu, Jamie Callan



November 2002 Proceedings of the eleventh international conference on Information and knowledge management

Full text available: mpdf(185.78 KB)

Additional Information: full citation, abstract, references, citings, index terms

Query-based sampling is a method of discovering the contents of a text database by submitting queries to a search engine and observing the documents returned. In prior research sampled documents were used to build resource descriptions for automatic database selection, and to build a centralized sample database for query expansion and result merging. An unstated assumption was that the associated storage costs were acceptable. When sampled documents are long, storage costs can be large. This pape ...

Keywords: distributed information retrieval, document pruning

40 The use of phrases and structured queries in information retrieval

W. Bruce Croft, Howard R. Turtle, David D. Lewis

September 1991 Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval

Full text available: pdf(1.35 MB)

Additional Information: full citation, references, citings, index terms

Results 21 - 40 of 200

Result page: <u>previous 1</u> **2** <u>3</u> <u>4</u> <u>5</u> <u>6</u> <u>7</u> <u>8</u> <u>9</u> <u>10</u> <u>next</u>

The ACM Portal is published by the Association for Computing Machinery. Copyright @ 2005 ACM, Inc. Terms of Usage Privacy Policy Code of Ethics Contact Us

Useful downloads: Adobe Acrobat QuickTime Windows Media Player Real Player



Subscribe (Full Service) Register (Limited Service, Free) Login

Search: 

The ACM Digital Library 
The Guide

- Carac

# THE ACH DIGITAL LIBRARY

Feedback Report a problem Satisfaction survey

## Concept based query expansion

**Full text** 

**Pdf** (1.05 MB)

Source

Annual ACM Conference on Research and Development in Information Retrieval <u>archive</u> Proceedings of the 16th annual international ACM SIGIR conference on Research and

development in information retrieval table of contents

Pittsburgh, Pennsylvania, United States

Pages: 160 - 169 Year of Publication: 1993 ISBN:0-89791-605-0

**Authors** 

Yonggang Qiu Hans-Peter Frei

**Sponsor** 

SIGIR: ACM Special Interest Group on Information Retrieval

Publisher ACM Press New York, NY, USA

Additional Information: abstract references citings index terms collaborative colleagues peer to peer

**Tools and Actions:** 

Discussions Find similar Articles Review this Article

Save this Article to a Binder Display Formats: BibTex EndNote

**DOI Bookmark:** 

Use this link to bookmark this Article: http://doi.acm.org/10.1145/160688.160713

What is a DOI?

#### **↑ ABSTRACT**

Query expansion methods have been studied for a long time - with debatable success in many instances. In this paper we present a probabilistic query expansion model based on a similarity thesaurus which was constructed automatically. A similarity thesaurus reflects domain knowledge about the particular collection from which it is constructed. We address the two important issues with query expansion: the selection and the weighting of additional search terms. In contrast to earlier methods, our queries are expanded by adding those terms that are most similar to the concept of the query, rather than selecting terms that are similar to the query terms. Our experiments show that this kind of query expansion results in a notable improvement in the retrieval effectiveness when measured using both recall-precision and usefulness.

## **↑ REFERENCES**

Note: OCR errors may be found in this Reference List extracted from the full text article. ACM has opted to expose the complete List rather than only correct and linked references.

Cro 90 C. J. Crouch, An approach to the automatic construction of global thesauri, Information Processing and Management: an International Journal, v.26 n.5, p.629-640, 1990

Cro 92 <u>Carolyn J. Crouch , Bokyung Yang, Experiments in automatic statistical thesaurus</u> construction, Proceedings of the 15th annual international ACM SIGIR conference on Research and

- development in information retrieval, p.77-88, June 21-24, 1992, Copenhagen, Denmark
- Dee 90 Deerwester, S., Dumais, S.T., Fumas, G.W., Landauer, T.K., Harshman R., Indexing by latent semantic analysis, J. of the ASIS, 41 (6): 391-407, 1990.
- Ekm 92 <u>F. Cuna Ekmekcioglu</u>, Alexander M. Robertson, Peter Willett, Effectiveness of query expansion in ranked-output document retrieval systems, Journal of Information Science, v.18 n.2, p.139-147, 1992
- Fre 91 H. P. Frei, P. Schäuble, Determining the effectiveness of retrieval algorithms, Information Processing and Management: an International Journal, v.27 n.2-3, p.153-164, Feb./March 1991
- Gre 92 Gregory Grefenstette, Use of syntactic context to produce term association lists for text retrieval, Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval, p.89-97, June 21-24, 1992, Copenhagen, Denmark
- Han 92 <u>Micheline Hancock-Beaulieu, Query expansion: advances in research in online catalogues, Journal of Information Science, v.18 n.2, p.99-103, 1992</u>
- Har92 Donna Harman, Relevance feedback revisited, Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval, p.1-10, June 21-24, 1992, Copenhagen, Denmark
- Les 69 Lesk, M.E., Word-word association in document retrieval systems, American Documentation, 20(1): 27-38, 1969.
- Min 72 Minker, J., Wilson, G.A., Zirnmerman, B.I-I., An evaluation of query expansion by the addition of clustered terms for a document retrieval system, Information Storage and Retrieval, 8(6): 329-48, 1972.
- Pea 91 Peat, H.J., Willett, P., The limitations of term co-occurrence data for query expansion in document retrieval systems, J. of the ASIS, 42(5): 378-83, 1991.
- Qiu 92 Qiu, Y., ISIR: an integrated system for information retrieval, Proc. 14th 1R Colloquium, British Computer Society, Lancaster, 1992.
- Rug 92 Gerda Ruge, Experiment on linguistically-based term associations, Information Processing and Management: an International Journal, v.28 n.3, p.317-332, 1992
- Sal 71 Salton, G., Experiments in automatic thesaurus construction for information retrieval, Information Processing 71, 1: 115-123, 1971.
- Sal 75 Salton, G., Yang, C.S., Yu, C.T., A theory of term importance in automatic text analysis, J. of the ASIS, 26(1): 33-44, 1975.
- Sal 80 Salton, G., Automatic term class construction using relevance-a summary of work in automatic pseudoclassification, Information Processing & Management, 16(1): 1-15, 1980.
- Sal 88 Gerard Salton, Christopher Buckley, Term-weighting approaches in automatic text retrieval, Information Processing and Management: an International Journal, v.24 n.5, p.513-523, 1988
- Sal 90 Salton, G., Buckley, C.: Improving Retrieval Performance by Relevance Feedback. J. of the ASIS, 41(4): 288-297, 1990.

- Sch 92 Sch~iuble, P., Knaus, D., The various roles of information structures, 16. Jahrestagung der Gesellschaft ffir Klassifikation, Dortmund, 1992.
- Sme 83 Smeaton, A.F., van Rijsbergen, C.J., The retrieval effects of query expansion on a feedback document retrieval system, The Computer Journal, 26(3): 239-46, 1983.
- Spa 71 Sparck-Jones, K., Barber, E.B., What makes an automatic keyword classification effective? J. of the ASIS, 18." 166-175, 1971.
- Spa 91 Karen Sparck Jones, Notes and references on early automatic classification work, ACM SIGIR Forum, v.25 n.1, p.10-17, Spring 1991
- Wad 88 Stephe J. Wade , Peter Willett, INSTRUCT: a teaching package for experimental methods in information retrieval. Part 111. Browsing, clustering and query, Program, v.22 n.1, p.44-61, January 1988
- Won 87 S. K.M. Wong, W. Ziarko, V. V. Raghavan, P. C.N. Wong, On modeling of information retrieval concepts in vector spaces, ACM Transactions on Database Systems (TODS), v.12 n.2, p.299-321, June 1987

#### ↑ CITINGS 37

<u>Justin Picard, Finding content-bearing terms using term similarities, Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics, June 08-12, 1999, Bergen, Norway</u>

David Hull, Improving text retrieval for the routing problem using latent semantic indexing, Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, p.282-291, July 03-06, 1994, Dublin, Ireland

<u>Páraic Sheridan</u>, <u>Jean Paul Ballerini</u>, <u>Experiments in multilingual information retrieval using the SPIDER system</u>, <u>Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval</u>, <u>p.58-65</u>, <u>August 18-22</u>, <u>1996</u>, <u>Zurich</u>, <u>Switzerland</u>

Hang Cui, Ji-Rong Wen, Jian-Yun Nie, Wei-Ying Ma, Probabilistic query expansion using query logs, Proceedings of the eleventh international conference on World Wide Web, May 07-11, 2002, Honolulu, Hawaii, USA

Jinxi Xu., W. Bruce Croft, Query expansion using local and global document analysis, Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval, p.4-11, August 18-22, 1996, Zurich, Switzerland

<u>Jian-Yun Nie, Query expansion and query translation as logical inference, Journal of the American Society for Information Science and Technology, v.54 n.4, p.335-346, February 15, 2003</u>

Rila Mandala, Takenobu Tokunaga, Hozumi Tanaka, Complementing WordNet with Roget's and corpus-based thesauri for information retrieval, Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics, June 08-12, 1999, Bergen, Norway

F. Crestani, C. J. van Rijsbergen, Probability kinematics in information retrieval, Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval, p.291-299, July 09-13, 1995, Seattle, Washington, United States

Susan Gauch, Jianying Wang, A corpus analysis approach for automatic query expansion,

<u>Proceedings of the sixth international conference on Information and knowledge management, p.278-284, November 10-14, 1997, Las Vegas, Nevada, United States</u>

P. D. Bruza, D. Song, Inferring query models by computing information flow, Proceedings of the eleventh international conference on Information and knowledge management, November 04-09, 2002, McLean, Virginia, USA

Hideo Joho, Claire Coverson, Mark Sanderson, Micheline Beaulieu, Hierarchical presentation of expansion terms, Proceedings of the 17th symposium on Proceedings of the 2002 ACM symposium on applied computing, March 11-14, 2002, Madrid, Spain

Rila Mandala, Takenobu Tokunaga, Hozumi Tanaka, Combining multiple evidence from different types of thesaurus for query expansion, Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, p.191-197, August 15-19, 1999, Berkeley, California, United States

Amit Singhal, Fernando Pereira, Document expansion for speech retrieval, Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, p.34-41, August 15-19, 1999, Berkeley, California, United States

Bienvenido Vélez , Ron Weiss , Mark A. Sheldon , David K. Gifford, Fast and effective query refinement, ACM SIGIR Forum, v.31 n.SI, p.6-15

Yuen-Hsien Tseng, Automatic thesaurus generation for Chinese documents, Journal of the American Society for Information Science and Technology, v.53 n.13, p.1130-1138, November 2002

Chien-Chung Huang, Shui-Lung Chuang, Lee-Feng Chien, Liveclassifier: creating hierarchical text classifiers through web corpora, Proceedings of the 13th conference on World Wide Web, May 17-20, 2004, New York, NY, USA

Susan Gauch, Jianying Wang, Satya Mahesh Rachakonda, A corpus analysis approach for automatic query expansion and its extension to multiple databases, ACM Transactions on Information Systems (TOIS), v.17 n.3, p.250-269, July 1999

Young Mee Chung, Jae Yun Lee, Optimization of some factors affecting the performance of query expansion, Information Processing and Management: an International Journal, v.40 n.6, p.891-917, November 2004

Bodo Billerbeck, Justin Zobel, Questioning query expansion: an examination of behaviour and parameters, Proceedings of the fifteenth conference on Australasian database, p.69-76, January 01, 2004, Dunedin, New Zealand

James Caverlee , Ling Liu , Daniel Rocco, Discovering and ranking web services with BASIL: a personalized approach with biased focus, Proceedings of the 2nd international conference on Service oriented computing, November 15-19, 2004, New York, NY, USA

<u>Paul McNamee</u>, <u>James Mayfield</u>, <u>Christine Piatko</u>, <u>Haircut:</u> a system for multilingual text retrieval in java, The Journal of Computing in Small Colleges, v.17 n.3, p.8-22, February 2002

Reiner Kraft , Jason Zien, Mining anchor text for query refinement, Proceedings of the 13th conference on World Wide Web, May 17-20, 2004, New York, NY, USA

Paul McNamee , James Mayfield, Comparing cross-language query expansion techniques by degrading translation resources, Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, August 11-15, 2002, Tampere, Finland

Farid Meziane, Yacine Rezgui, A document management methodology based on similarity contents, Information Sciences—Informatics and Computer Science: An International Journal, v.158 n.1, p.15-36, January 2004

Nega Alemayehu, Analysis of performance variation using query expansion, Journal of the American Society for Information Science and Technology, v.54 n.5, p.379-391, March 2003

<u>Luis M. de Campos</u>, <u>Juan M. Fernández-Luna</u>, <u>Juan F. Huete, Improving the efficiency of the Bayesian network retrieval model by reducing relationships between terms, International Journal of <u>Uncertainty, Fuzziness and Knowledge-Based Systems</u>, v.11 n.Supplement, p.101-116, September 2003</u>

Shuang Liu, Fang Liu, Clement Yu, Weiyi Meng, An effective approach to document retrieval via utilizing WordNet and recognizing phrases, Proceedings of the 27th annual international conference on Research and development in information retrieval, July 25-29, 2004, Sheffield, United Kingdom

Alberto Lavelli, Fabrizio Sebastiani, Roberto Zanoli, Distributional term representations: an experimental comparison, Proceedings of the Thirteenth ACM conference on Information and knowledge management, November 08-13, 2004, Washington, D.C., USA

Maria L. Silveira, Berthier Ribeiro-Neto, Concept-based ranking: a case study in the juridical domain, Information Processing and Management: an International Journal, v.40 n.5, p.791-805, September 2004

Makoto Nakashima, Keizo Sato, Yanhua Qu, Tetsuro Ito, Browsing-based conceptual information retrieval incorporating dictionary term relations, keyword association, and a user's interest, Journal of the American Society for Information Science and Technology, v.54 n.1, p.16-28, January 2003

F. Crestani , C. J. van Rijsbergen, A study of probability kinematics in information retrieval, ACM Transactions on Information Systems (TOIS), v.16 n.3, p.225-255, July 1998

Jinxi Xu , W. Bruce Croft, Improving the effectiveness of information retrieval with local context analysis, ACM Transactions on Information Systems (TOIS), v.18 n.1, p.79-112, Jan. 2000

Christophe Brouard, Jian-Yun Nie, Relevance as resonance: a new theoretical perspective and a practical utilization in information filtering, Information Processing and Management: an International Journal, v.40 n.1, p.1-19, January 2004

Claudio Carpineto, Giovanni Romano, Vittorio Giannini, Improving retrieval feedback with multiple term-ranking function combination, ACM Transactions on Information Systems (TOIS), v.20 n.3, p.259-290, July 2002

<u>Hinrich Schütze, Automatic word sense discrimination, Computational Linguistics, v.24 n.1, March 1998</u>

Raymond Y. K. Lau, Context-sensitive text mining and belief revision for intelligent information retrieval on the web, Web Intelligence and Agent System, v.1 n.3-4, p.151-172, March 2003

Fabio Crestani, Mounia Lalmas, Logic and uncertainty in information retrieval, Lectures on information retrieval, Springer-Verlag New York, Inc., New York, NY, 2001

#### **↑ INDEX TERMS**

#### **Primary Classification:**

H. Information Systems

+ H.3 INFORMATION STORAGE AND RETRIEVAL

H.3.3 Information Search and Retrieval

Subjects: Query formulation

#### Additional Classification:

H. Information Systems

+.3 INFORMATION STORAGE AND RETRIEVAL

H.3.1 Content Analysis and Indexing

Subjects: Indexing methods; Thesauruses

#### **General Terms:**

Algorithms, Experimentation, Languages

## ♠ Collaborative Colleagues:

Hans-Peter Frei: M. Bärtschi

Donna Harman S. Meienberg Yonggang Qiu Peter Schäuble Peter Schaüble

Gabriele Sonnenberger

D. Stieger Ross Wilkinson

Yonggang Qiu: Hans-Peter Frei

## ↑ Peer to Peer - Readers of this Article have also read:

M4: a metamodel for data preprocessing

Proceedings of the 4th ACM international workshop on Data warehousing and OLAP Anca Vaduva , Jörg-Uwe Kietz , Regina Zücker

- <u>Data structures for quadtree approximation and compression</u>
   <u>Communications of the ACM</u> 28, 9
   Hanan Samet
- A hierarchical single-key-lock access control using the Chinese remainder theorem
   Proceedings of the 1992 ACM/SIGAPP Symposium on Applied computing
   Kim S. Lee , Huizhu Lu , D. D. Fisher
- The GemStone object database management system
   Communications of the ACM 34, 10
   Paul Butterworth , Allen Otis , Jacob Stein
- Putting innovation to work: adoption strategies for multimedia communication systems
   Communications of the ACM 34, 12
   Ellen Francik, Susan Ehrlich Rudman, Donna Cooper, Stephen Levine

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2005 ACM, Inc.

<u>Terms of Usage Privacy Policy Code of Ethics Contact Us</u>

Useful downloads: Adobe Acrobat QuickTime Windows Media Player Real Player

THIS PAGE BLANK (USP10,



Subscribe (Full Service) Register (Limited Service, Free) Login

Search: The ACM Digital Library The Guide

(cluster < near/4> concepts) frequency occurrences



# THE ACM DIGITAL LIBRARY

Feedback Report a problem Satisfaction survey

Terms used cluster near/4 concepts frequency occurrences

Found 26,328 of 154,226

Sort results

þγ Display

results

relevance

expanded form

Save results to a Binder Search Tips Open results in a new

Try an Advanced Search Try this search in The ACM Guide

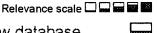
Results 101 - 120 of 200

window

Result page: previous 1 2 3 4 5 6 7 8 9 10

Best 200 shown

101 A methodology for supporting existing CODASYL databases with new database



machines Jayanta Banerjee, David K. Hsiao

January 1978 Proceedings of the 1978 annual conference - Volume 2

Full text available: pdf(1.05 MB)

Additional Information: full citation, abstract, references, citings, index terms

In this paper, an attempt is made to show that conventional database management system software, in particular those of CODASYL type, can be effectively replaced by database machines with good performance. The replacement of CODASYL system software involves two main steps: (i) In order to preserve the notions of CODASYL records, sets,areas, and others, we need a methodology for database transformation so that an existing CODASYL database may be transformed into suita ...

Keywords: CODASYL data model, DBC, Database machines, Database management systems, Database transformation, Network data model, Query translation, Relative performance.

102 Multikey access methods based on term discrimination and signature clustering J. W. Chang, J. H. Lee, Y. J. Lee



May 1989 ACM SIGIR Forum, Proceedings of the 12th annual international ACM SIGIR conference on Research and development in information retrieval, Volume 23 Issue 1-2

Full text available: pdf(963.83 KB)

Additional Information: full citation, abstract, references, citings, index

In order to improve the two-level signature file method designed by Sacks-Davis et al. [20], we propose new multikey access methods based on term discrimination and signature clustering. By term discrimination, we create separate, efficient access methods for the terms frequently used in user queries. We in addition cluster similar signatures by means of these terms so that we may achieve good performance on retrieval. Meanwhile we provide the space-time analysis of the proposed methods and ...

103 Clustering: Learning to cluster web search results

Hua-Jun Zeng, Qi-Cai He, Zheng Chen, Wei-Ying Ma, Jinwen Ma

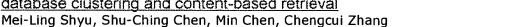
July 2004 Proceedings of the 27th annual international conference on Research and development in information retrieval

Full text available: pdf(210.38 KB) Additional Information: full citation, abstract, references, citings, index terms

Organizing Web search results into clusters facilitates users' quick browsing through search results. Traditional clustering techniques are inadequate since they don't generate clusters with highly readable names. In this paper, we reformalize the clustering problem as a salient phrase ranking problem. Given a query and the ranked list of documents (typically a list of titles and snippets) returned by a certain Web search engine, our method first extracts and ranks salient phrases as candidate c ...

Keywords: document clustering, regression analysis, search result organization

104 Content-based retrieval for multimedia databases: A unified framework for image database clustering and content-based retrieval



November 2004 Proceedings of the 2nd ACM international workshop on Multimedia databases

Full text available: pdf(291.68 KB) Additional Information: full citation, abstract, references, index terms

With the proliferation of image data, the need to search and retrieve images efficiently and accurately from a large image database or a collection of image databases has drastically increased. To address such a demand, a unified framework called <i>Markov Model Mediators</i> (MMMs) is proposed in this paper to facilitate conceptual database clustering and to improve the query processing performance by analyzing the summarized knowledge. The unique characteristics of MMMs are that it ...

**Keywords:** Markov model mediators (MMMs), content-based image retrieval (CBIR), image database clustering

105 Research track: Information-theoretic co-clustering

Inderjit S. Dhillon, Subramanyam Mallela, Dharmendra S. Modha

August 2003 Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining

Full text available: pdf(1.96 MB)

Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index</u> <u>terms</u>

Two-dimensional contingency or co-occurrence tables arise frequently in important applications such as text, web-log and market-basket data analysis. A basic problem in contingency table analysis is co-clustering: simultaneous clustering of the rows and columns. A novel theoretical formulation views the contingency table as an empirical joint probability distribution of two discrete random variables and poses the co-clustering problem as an optimization problem in information theory

**Keywords**: co-clustering, information theory, mutual information

106 A hybrid handover protocol for local area wireless ATM networks

Chai-Keong Toh

December 1996 Mobile Networks and Applications, Volume 1 Issue 3

Full text available: pdf(960.44 KB)

Additional Information: full citation, abstract, references, citings, index terms

While handovers of voice calls in a wide area mobile environment are well understood, handovers of multi-media traffic in a local area mobile environment is still in its early stage of investigation. Unlike the public wireless networks, handovers for multi-media Wireless LANS (WLANS) have special requirements. In this paper, the problems and challenges faced in a multi-media WLAN environment are outlined and a multi-tier wireless cell





clustering architecture is introduced. Design issues for ...

107 Cut as a querving unit for WWW, Netnews, e-mail

Keishi Tajima, Yoshiaki Mizuuchi, Masatsugu Kitagawa, Katsumi Tanaka

May 1998 Proceedings of the ninth ACM conference on Hypertext and hypermedia: links, objects, time and space---structure in hypermedia systems: links, objects, time and space---structure in hypermedia systems

Full text available: pdf(1.23 MB)

Additional Information: full citation, references, citings, index terms

108 Text categorization: Text categorization by boosting automatically extracted concepts Lijuan Cai, Thomas Hofmann



July 2003 Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval

Full text available: pdf(237.53 KB)

Additional Information: full citation, abstract, references, citings, index terms

Term-based representations of documents have found wide-spread use in information retrieval. However, one of the main shortcomings of such methods is that they largely disregard lexical semantics and, as a consequence, are not sufficiently robust with respect to variations in word usage. In this paper we investigate the use of concept-based document representations to supplement word- or phrase-based features. The utilized concepts are automatically extracted from documents via probabilistic late ...

Keywords: boosting, classification, concept extraction, document categorization, lexical semantics, machine learning

109 Industrial/government track: Visualizing concept drift



August 2003 Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining

Full text available: pdf(580.02 KB) Additional Information: full citation, abstract, references, index terms

We describe a visualization technique that uses brushed, parallel histograms to aid in understanding concept drift in multidimensional problem spaces. This technique illustrates the relationship between changes in distributions of multiple antecedent feature values and the outcome distribution. We can also observe effects on the relative utilization of predictive rules. Our parallel histogram technique solves the over-plotting difficulty of parallel coordinate graphs and the difficulty of compar ...

Keywords: brushing, concept drift, parallel coordinate graph, parallel histogram, visualization

110 Full text indexing based on lexical relations an application: software libraries Y. S. Maarek, F. Z. Smadja



May 1989 ACM SIGIR Forum, Proceedings of the 12th annual international ACM SIGIR conference on Research and development in information retrieval, Volume 23 Issue 1-2

Full text available: pdf(970.81 KB)

Additional Information: full citation, abstract, references, citings, index terms

In contrast to other kinds of libraries, software libraries need to be conceptually organized. When looking for a component, the main concern of users is the functionality of the desired component; implementation details are secondary. Software reuse would be

enhanced with conceptually organized large libraries of software components. In this paper, we present GURU, a tool that allows automatical building of such large software libraries from documented software components. We focus here on ...

111 Performance and implications of semantic indexing in a distributed environment Conrad T. K. Chang, Bruce R. Schatz



November 1999 Proceedings of the eighth international conference on Information and knowledge management

Full text available: pdf(985.41 KB)

Additional Information: full citation, abstract, references, citings, index

A research prototype is presented for semantic indexing and retrieval in Information Retrieval. The prototype is motivated by a desire to provide a more efficient and effective information retrieval system compared to the current state of the art. An overview of the Interspace architecture layers is discussed. An object model supporting semantic operations is developed. The model contains a rich set of classes and relationships of the data for the semantic indexing module. The basis of our ...

Keywords: concept space, distributed computing, information retrieval, semantic indexing

112 Compression, information theory, and grammars: a unified approach Abraham Bookstein, Shmuel T. Klein January 1990 ACM Transactions on Information Systems (TOIS), Volume 8 Issue 1



Full text available: pdf(1.80 MB)

Additional Information: full citation, abstract, references, citings, index terms

Text compression is of considerable theoretical and practical interest. It is, for example, becoming increasingly important for satisfying the requirements of fitting a large database onto a single CD-ROM. Many of the compression techniques discussed in the literature are model based. We here propose the notion of a formal grammar as a flexible model of text generation that encompasses most of the models offered before as well as, in principle, extending the possibility of compression to a ...

113 Research track: Fragments of order

Aristides Gionis, Teija Kujala, Heikki Mannila

August 2003 Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining

Full text available: pdf(135.83 KB) Additional Information: full citation, abstract, references, index terms

High-dimensional collections of 0--1 data occur in many applications. The attributes in such data sets are typically considered to be unordered. However, in many cases there is a natural total or partial order ≺ underlying the variables of the data set. Examples of variables for which such orders exist include terms in documents, courses in enrollment data, and paleontological sites in fossil data collections. The observations in such applications are flat, unordered sets; however, the data s ...

Keywords: consecutive ones property, discovering hidden orderings, novel data mining algorithms, spectral analysis of data

114 Structural disambiguation based on reliable estimation of strength of association Haodong Wu, Eduardo de Paiva Alves, Teiji Furugori August 1998



Full text available: pdf(574.67 KB)

Additional Information:



full citation, abstract, references

This paper proposes a new class-based method to estimate the strength of association in word co-occurrence for the purpose of structural disambiguation. To deal with sparseness of data, we use a conceptual dictionary as the source for acquiring upper classes of the words related in the co-occurrence, and then use t-scores to determine a pair of classes to be employed for calculating the strength of association. We have applied our method to determining dependency relations in Japanese and prepos ...

115 THESUS: Organizing Web document collections based on link semantics Maria Halkidi, Benjamin Nguyen, Iraklis Varlamis, Michalis Vazirgiannis November 2003 The VLDB Journal — The International Journal on Very Large Data Bases, Volume 12 Issue 4

Full text available: pdf(262.85 KB) Additional Information: full citation, abstract, citings, index terms

The requirements for effective search and management of the WWW are stronger than ever. Currently Web documents are classified based on their content not taking into account the fact that these documents are connected to each other by links. We claim that a page's classification is enriched by the detection of its incoming links' semantics. This would enable effective browsing and enhance the validity of search results in the WWW context. Another aspect that is underaddressed and str ...

Keywords: Document clustering, Link analysis, Link management, Semantics, Similarity measure, World Wide Web

116 Visualization of search results: a comparative evaluation of text. 2D, and 3D interfaces



Marc M. Sebrechts, John V. Cugini, Sharon J. Laskowski, Joanna Vasilakis, Michael S. Miller August 1999 Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval

Full text available: Top pdf(1.68 MB)

Additional Information: full citation, references, citings, index terms

Keywords: 2D, 3D, evaluation, information visualization, interface design, text

117 Session: Term extraction + term clustering: an integrated platform for computer-aided terminology



Didier Bourigault, Christian Jacquemin

June 1999 Proceedings of the ninth conference on European chapter of the **Association for Computational Linguistics** 

Full text available: pdf(749.27 KB) Publisher Site

Additional Information: full citation, abstract, references, citings

A novel technique for automatic thesaurus construction is proposed. It is based on the complementary use of two tools: (1) a Term Extraction tool that acquires term candidates from tagged corpora through a shallow grammar of noun phrases, and (2) a Term Clustering tool that groups syntactic variants (insertions). Experiments performed on corpora in three technical domains yield clusters of term candidates with precision rates between 93% and 98%.

118 "Is this document relevant?...probably": a survey of probabilistic models in information retrieval



Fabio Crestani, Mounia Lalmas, Cornelis J. Van Rijsbergen, Iain Campbell

## December 1998 ACM Computing Surveys (CSUR), Volume 30 Issue 4

Additional Information: full citation, abstract, references, citings, index Full text available: pdf(265.85 KB) terms, review

This article surveys probablistic approaches to modeling information retrieval. The basic concepts of probabilistic approaches to information retrieval are outlined and the principles and assumptions upon which the approaches are based are presented. The various models proposed in the development of IR are described, classified, and compared using a common formalism. New approaches that constitute the basis of future research are described.

**Keywords:** information retrieval, probabilistic indexing, probabilistic modeling, probabilistic retrieval, uncertain inference modeling

119 Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies



Soumen Chakrabarti, Byron Dom, Rakesh Agrawal, Prabhakar Raghavan August 1998 The VLDB Journal — The International Journal on Very Large Data Bases, Volume 7 Issue 3

Full text available: pdf(281.37 KB) Additional Information: full citation, abstract, citings, index terms

We explore how to organize large text databases hierarchically by topic to aid better searching, browsing and filtering. Many corpora, such as internet directories, digital libraries, and patent databases are manually organized into topic hierarchies, also called taxonomies. Similar to indices for relational data, taxonomies make search and access more efficient. However, the exponential growth in the volume of on-line textual information makes it nearly impossible to maintain such taxono ...

# 120 Graphical table of contents



April 1996 Proceedings of the first ACM international conference on Digital libraries

Full text available: pdf(978.24 KB) Additional Information: full citation, references, citings, index terms

Results 101 - 120 of 200

Result page: previous 1 2 3 4 5 6 7 8 9 10 next

The ACM Portal is published by the Association for Computing Machinery. Copyright @ 2005 ACM, Inc. Terms of Usage Privacy Policy Code of Ethics Contact Us

Useful downloads: Adobe Acrobat QuickTime Windows Media Player Real Player



Home | Login | Logout | Access Information | Alerts |

Welcome United States Patent and Trademark Office

BROWSE

SEARCH

IEEE XPLORE GUIDE

'HIS PAGE BLANK (USPTO)



Home | Login | Logout | Access Information | Alerts |

Welcome United States Patent and Trademark Office

BROWSE

SEARCH

IEEE XPLORE GUIDE

You have entered an invalid search.

Please verify your search criteria and try again.

If you believe you have reached this page in error, please complete the online Technical Support Fc onlinesupport@ieee.org.

Message #VP2005



Welcome to IEEE Xplore

computer science, and electronics.

... delivering full text access to the world's highest quality technical literature in electrical engineering,

1,157,693 documents online

Home | Login | Logout | Access Information | Aierts |

#### Welcome United States Patent and Trademark Office

BROWSE

SEARCH

IEEE XPLORE GUIDE

## Browse

- » Journals & Magazines
- » Conference Proceedings
- » Standards

## Basic Search

(((cluster <near/4> concepts)

(All Fields)

- » Advanced Search
- » Author Search
- » CrossRef Search

Content Updates Browse the latest update to see recently added content.

» Latest Content Update

Aleris

Register and access your tables of contents alerts.

» Visit Alerts

**Top 100 Documents** Find out the most accessed documents for the month.

» View Top 100



Cookies Enabled

indexed by # inspec Contact Us Privacy &:

@ Copyright 2005 IEEE -



Heme | Login | Logout | Access Information | Alerts

Welcome United States Patent and Trademark Office

BROWSE

SEARCH

IEEE XPLORE GUIDE

**OPTION 1** 

Quick Find an Author:

Enter a name to locate articles written by that author.

gallivan  $\Sigma$ 

Example: Enter Lockett S to obtain a list of authors with the last name Lockett and the first initial S.

**OPTION 2** 

Browse alphabetically

Select a letter from the list.

<u>A B C D E F G H I J K L M N O P Q R S T U V W X Y Z</u>

Select a name to view articles written by that author

Gallivan J.

Gallivan K.

<u>Gallivan</u>

Gallivan M. A.

<u>Gallivan</u>

Gallivan M. Gallivanoni A.

indexed by # inspec Help Contact Us Privacy &

© Copyright 2005 IEEE



Web Images Groups News Froogle Local more »

(cluster concepts) frequency occurrences

Search

Advanced Search Preferences

## Web

Results 1 - 10 of about 72,200 for (cluster concepts) frequency occurrences. (0.16 seconds)

Tip: Save time by hitting the return key instead of clicking on "search"

## DiploFoundation - The Emerging Language of Information Society

... We present the **frequency** distributions of the key **concepts occurrences** in ... multidimensional scaling and **cluster** analysis - were performed in order to ... www.diplomacy.edu/IS/Language/html/concepts.htm - 56k - <u>Cached</u> - <u>Similar pages</u>

## DiploFoundation - The Emerging Language of Information Society

... Hereby we present the **frequency** distributions of keywords **occurrences**: ... As in the previous analysis of key **concepts frequency** distributions, ...

www.diplomacy.edu/IS/Language/html/words.htm - 53k - Cached - Similar pages

## Diagnosis as defined by the clinician

... about individual occurrences or exacerbations. Cluster (0..1) optional ...

Frequency of occurrences, The frequency of occurrences of the problem ...
oceaninformatics.biz/archetypes/ openEHR-EVALUATION.problem-diagnosis.v1.html - 19k - May 16, 2005 - Cached - Similar pages

## archetype openEHR-EHR-EVALUATION.problem.v1 concept [at0000 ...

... unordered) matches { CLUSTER[at0027] occurrences matches {0. ... text = <"Frequency of occurrences"> description = <"The frequency of occurrences of the ... oceaninformatics.biz/archetypes/ ADL/openEHR-EHR-EVALUATION.problem.v1.adl - 13k - Cached - Similar pages

# [PDF] Concept Decompositions for Large Sparse Text Data Using Clustering

File Format: PDF/Adobe Acrobat - View as HTML

... Lower Frequency Bound: 0.2 % of the documents with low frequency words ... Along with each cluster, we associate a concept ...

bayrak.ualr.edu/symsel/hemant/ docs/Concept%20Decompositions.pdf - Similar pages

## Sample Studies Using WordStat

... This procedure identified 12 clusters of concepts, representing areas of significant ... Word co-occurrences were identified and then cluster analyzed, ... www.kovcomp.co.uk/wordstat/studies.html - 28k - Cached - Similar pages

## греп Information Organization and Retrieval

File Format: Microsoft Powerpoint 97 - View as HTML

... The product of the **frequency** of words (f) and their rank (r) is ... **Cluster** sets of documents into general "themes", like a table of contents ... www.sims.berkeley.edu/courses/ is202/f00/lectures/Lecture6\_202.ppt - <u>Similar pages</u>

# <u>Selecting co-occurrences from MRCOC for integration into MRREL ...</u> ... Ignore co-occurrences in concepts whose total frequency of co-occurrence is

... eventually recompute the co-occurrence of clusters instead of concepts). ... umlsinfo.nlm.nih.gov/coocrels/COCRELS.TXT - 17k - Cached - Similar pages

# [PDF] Chapter HDDI ä: Hierarchical Distributed Dynamic Indexing

File Formal: PDF/Adobe Acrobat - View as HTML

... the frequencies of co-occurrences of concept pairs among all items in the set.

... clusters of concepts œ subtopic regions of semantic locality. ... www.cse.lehigh.edu/~billp/pubs/HDDIFinalChapter.pdf - Similar pages

гррт www.cs.cornell.edu/courses/cs430/2004fa/slides/lec... File Format: Microsoft Powerpoint 97 - View as HTML ... fi is the number of occurrences of term i in the corpus. ... stage is to group similar terms together into concepts. This is done by cluster analysis. ... Similar pages

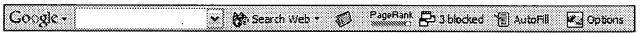
Goooooooogle >

Result Page:

1 2 3 4 5 6 7 8 9 10

Next

Free! Get the Google Toolbar. Download Now - About Toolbar



(cluster concepts) frequency occurre Search

Search within results | Language Tools | Search Tips | Dissatisfied? Help us improve

Google Home - Advertising Programs - Business Solutions - About Google

©2005 Google

Google

Web Images Groups News Froogle Local more »

(cluster concepts) frequency occurrences

Search

Advanced Search Preferences

## Web

Results 11 - 20 of about 72,200 for (cluster concepts) frequency occurrences. (0.04 seconds)

## [PDF] Tim Oates, David Jensen and Paul R. Cohen Introduction

File Format: PDF/Adobe Acrobat - View as HTML

... tant activities: **clustering** sets of related events and, predicting events that will occur in ... pectedly frequent or infrequent co-**occurrences** of pat- ... www-eksl.cs.umass.edu/papers/oates-aaai98-time.pdf - Similar pages

## грет Information Organization and Retrieval

File Format: Microsoft Powerpoint 97 - View as HTML

... Word frequencies have a Zipf distribution. Word co-occurrences exhibit ... Clustering Multi-Dimensional Document Space. Wise et al., 1995. Concept ...

www.sims.berkeley.edu/courses/is202/f01/Lectures/Lecture-9-13-01.ppt - Similar pages

## [PDF] SenseClusters - Finding Clusters that Represent Word Senses

File Format: PDF/Adobe Acrobat - View as HTML

... a cluster are contextually similar to each other, making it ... occurrences are bigrams that include the target word. In. effect co-occurrences localize ... www.d.umn.edu/~tpederse/ Pubs/naacl04-demo-senseclusters.pdf - Similar pages

## Machine Learning

- ... identify common clusters of problems and complaints on certain automobiles, etc.
- ... that both reflect the relative frequencies of occurrences of words, ...

www.statsoft.com/textbook/sttextmin.html - 22k - Cached - Similar pages

## **Indexing Methods**

... data such as the number of **occurrences** of the word in all documents, ... in a document then the more relevant it is - this is called Term **Frequency**. ... www.doc.ic.ac.uk/~nd/surprise\_\_97/ journal/vol4/mjc4/indexing.html - 10k - <u>Cached</u> - <u>Similar pages</u>

## [PDF] Detecting Emerging Concepts in Textual Data Mining

File Format: PDF/Adobe Acrobat - View as HTML

... We also compute the frequencies of co-occurrences of. concept pairs among all items in the ... We thus detect clusters of concepts within a knowledge ... www.cse.lehigh.edu/~billp/pubs/SIAMETD.pdf - Similar pages

## [PDF] A Clustering Method and Radius Tunin

File Format: PDF/Adobe Acrobat

... reflecting mainly the frequencies of character /cluster. occurrences. ... B.

Select the cluster with the largest Vohnne (Concept ...

doi.leeecomputersociety.org/10.1109/ICDAR.1995.601999 - Similar pages

## [PDF] Concept Extraction and Association from Cancer Literature

File Format: PDF/Adobe Acrobat - View as HTML

... concepts. VCGS (Vocabulary Cluster Generating System) was ... Convert the frequency of each unique token/document to a ...

lair.indiana.edu/research/lucas/papers/widm02.pdf - Similar pages

## грэт Chapter 5 Query Languages

File Format Microsoft Powerpoint 97 - View as HTML

... local cluster. basic concept. Expanding the query with terms correlated to the

... association clusters: co-occurrences of pairs of terms in documents ... nlg3.csie.ntu.edu.tw/courses/ IR/slides/Chapter\_5\_Query\_Operations.ppt - <u>Similar pages</u>

[PDF] The Clustering Power of Low Frequency Words in Academic Webs File Format: PDF/Adobe Acrobat - View as HTML

... The value of low **frequency** words for subject-based academic web site **clustering** is assessed. ... in terms of distinct word **occurrences** (Zipf, 1949). A ... www.scit.wlv.ac.uk/~cm1993/papers/ 2005\_Low\_**frequency**\_words\_in\_academic\_webs\_preprint.pdf - Similar pages



Result Page: <u>Previous 1 2 3 4 5 6 7 8 9 1011</u> Next

(cluster concepts) frequency occurre Search

Search within results | Language Tools | Search Tips

Google Home - Advertising Programs - Business Solutions - About Google

©2005 Google